

Lie geometry of 2×2 Markov matrices

Jeremy G. Sumner

School of Mathematics and Physics

University of Tasmania, Australia

ARC Research Fellow

email: jsumner@utas.edu.au

December 24, 2012

Abstract

In recent work discussing model choice for continuous-time Markov chains, we have argued that it is important that the Markov matrices that define the model are closed under matrix multiplication [6, 7]. The primary requirement is then that the associated set of rate matrices form a Lie algebra. For the generic case, this connection to Lie theory seems to have first been made by [3], with applications for specific models given in [1] and [2]. Here we take a different perspective: given a model that forms a Lie algebra, we apply existing Lie theory to gain additional insight into the geometry of the associated Markov matrices. In this short note, we present the simplest case possible of 2×2 Markov matrices. The main result is a novel decomposition of 2×2 Markov matrices that parameterises the general Markov model as a perturbation away from the binary-symmetric model. This alternative parameterisation provides a useful tool for visualising the binary-symmetric model as a submodel of the general Markov model.

keywords: Lie algebras, algebra, symmetry, Markov chains, phylogenetics

1 Results

Consider the set of real 2×2 Markov matrices

$$\left\{ \begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix} : a, b \in \mathbb{R} \right\},$$

and the subset of 2×2 “stochastic” Markov matrices

$$\left\{ \begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix} : 0 \leq a, b \leq 1 \in \mathbb{R} \right\}.$$

In models of phylogenetic molecular evolution (see for example [4]), this set provides the transition matrices for what is known as the “general Markov model” on two states. If we were to take the additional constraint $a = b$, the model would then be referred to as “binary-symmetric”.

Associated with these sets is the matrix group

$$\mathcal{G} := \left\{ \begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix} : a, b \in \mathbb{R}, a+b \neq 1 \right\}.$$

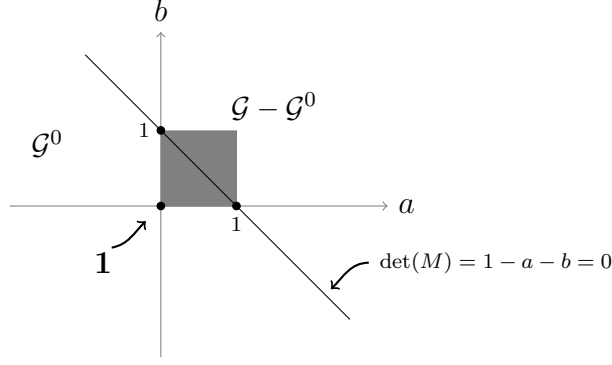


Figure 1: The group \mathcal{G} of invertible 2×2 Markov matrices of the form $\begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix}$ understood geometrically as a manifold in \mathbb{R}^2 . The gray area indicates the subset of “stochastic” Markov matrices. The line $\det(M) = 0$ indicates the boundary of the connected component to the identity, \mathcal{G}^0 .

We can geometrically understand \mathcal{G} by considering it as a manifold in \mathbb{R}^2 . This is illustrated in Figure 1.

By considering smooth paths $A(t) \in \mathcal{G}$, we can define the tangent space of this matrix group at the identity:

$$T_1(\mathcal{G}) = \{A'(0) : A(t) \in \mathcal{G} \text{ and } A(0) = \mathbf{1}\}.$$

As \mathcal{G} is a matrix group, it follows that $T_1(\mathcal{G})$ forms a *Lie algebra*. This means that for all $X, Y \in T_1(\mathcal{G})$ and $\lambda \in \mathbb{R}$, we have:

1. $X + \lambda Y \in T_1(\mathcal{G})$, ie. $T_1(\mathcal{G})$ is a vector space,
2. $[X, Y] := XY - YX \in T_1(\mathcal{G})$.

Consider two smooth functions $a(t)$ and $b(t)$ satisfying $a(t) + b(t) \neq 1$ for all t , and $a(0) = b(0) = 0$. Define

$$A(t) = \begin{pmatrix} 1 - a(t) & b(t) \\ a(t) & 1 - b(t) \end{pmatrix}.$$

Then, by construction, $A(t)$ is a smooth path in \mathcal{G} and $A'(0) \in T_1(\mathcal{G})$. If we define $L_1 := \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ and $L_2 := \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}$, we have $A'(0) = a'(0)L_1 + b'(0)L_2$, so $T_1(\mathcal{G}) = \langle L_1, L_2 \rangle_{\mathbb{R}}$ and $\{L_1, L_2\}$ is a basis for $T_1(\mathcal{G})$. It is straightforward to check that $[L_1, L_2] = L_1 - L_2$, so we conclude that $T_1(\mathcal{G})$ is indeed a Lie algebra.

Recall that a subgroup $H \leq G$ of a group is *normal* if $ghg^{-1} \in H$ for all $h \in H$ and $g \in G$. Also recall that the connected component to the identity G^0 is normal in G . In our case, this becomes:

Result 1. $\mathcal{G}^0 = \{M \in \mathcal{G} : \det(M) > 0\}$.

Proof. Consider $M = \begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix} = e^{Qt}$ where $Q := \begin{pmatrix} -\alpha & \beta \\ \alpha & -\beta \end{pmatrix}$ is a rate matrix (as would occur in a continuous-time formulation of a Markov process). Using the power series expansion of e^{Qt} , it is straightforward to show that, if $(a+b)t < 1$,

$$\alpha = \frac{-\log(1 - (a+b)t)}{1 + b/a}, \quad \beta = \frac{-\log(1 - (a+b)t)}{1 + a/b},$$

provides a solution to $M = e^{Qt}$. If we define the path $A(t) := e^{Qt}$, we have $A(0) = \mathbf{1}$ and $A(1) = M$. Thus, $M \in \mathcal{G}^0$ for all $a+b < 1$. On the other hand, if $a+b \geq 1$, there can be no path $B(t) \in \mathcal{G}$ with $B(0) = \mathbf{1}$ and $B(1) = M$ because we would have $\det(B(\tau)) = 0$ for some τ in the interval $(0, 1]$. \square

Corollary 1. $\mathcal{G}^0 = \left\{ e^Q : Q = \begin{pmatrix} -\alpha & \beta \\ \alpha & -\beta \end{pmatrix}; \alpha, \beta \in \mathbb{R} \right\}.$

Recall the homomorphism theorem for groups (see for example [5]), which ensures, for any group homomorphism $\rho : G \rightarrow G'$, that (i.) the image of ρ is a subgroup of G' , (ii.) the kernel K of ρ is normal in G , and (iii.) $G/K \cong G'$. To understand the set difference $\mathcal{G} - \mathcal{G}^0$, we notice that \mathcal{G}^0 is the kernel of the homomorphism,

$$\begin{aligned} \mathcal{G} &\rightarrow \{1, -1\} \cong \mathbb{Z}_2, \\ M &\mapsto \text{sgn}(\det(M)). \end{aligned}$$

The kernel of this homomorphism is \mathcal{G}^0 , thus $\mathcal{G}/\mathcal{G}^0 = \{\mathcal{G}^0, P\mathcal{G}^0\} \cong \mathbb{Z}_2$ for some $P \in \mathcal{G} - \mathcal{G}^0$. For reasons of symmetry, we reflect the identity $\mathbf{1}$ in the line $\det(M)$ and set $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, noting that $P^2 = \mathbf{1}$. As $\mathcal{G}/\mathcal{G}^0$ is a partition of \mathcal{G} , we see that $\mathcal{G} - \mathcal{G}^0 = P\mathcal{G}^0$ and $\mathcal{G} = \mathcal{G}^0 \cup P\mathcal{G}^0$.

Somewhat trivially:

Result 2. *As manifolds, $\mathcal{G}^0 \cong P\mathcal{G}^0$.*

Proof. Clearly,

$$\begin{aligned} P : \mathcal{G}^0 &\rightarrow P\mathcal{G}^0 \\ M &\mapsto PM, \end{aligned}$$

is a diffeomorphism because it maps continuous paths to continuous paths. \square

In particular, this means that:

Result 3. \mathcal{G}^0 is connected $\Leftrightarrow P\mathcal{G}^0$ is connected

Proof. No proof is required, but we give one regardless to illustrate. Consider the path $A(t) = e^{Q_2 t} e^{Q_1(1-t)} \in \mathcal{G}^0$ with $A(0) = M_1 := e^{Q_1}$ and $A(1) = M_2 := e^{Q_2}$. Now, $B(t) := PA(t)$ is a path in $P\mathcal{G}^0$ with $B(0) = PM_1$ and $B(1) = PM_2$. As any two points in $P\mathcal{G}^0$ can be written in this way, we are done. \square

Recall that the *center* $Z(G)$ of a group G is the set of all $g \in G$ such that $gh = hg$ for all $h \in G$. In our case, suppose that $N = \begin{pmatrix} 1-c & d \\ c & 1-d \end{pmatrix} \in Z(\mathcal{G})$. Setting $NM = MN$ implies:

$$\begin{aligned} \begin{pmatrix} 1-c & d \\ c & 1-d \end{pmatrix} \begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix} &= \begin{pmatrix} * & b(1-c-d) + d \\ a(1-d-c) + c & * \end{pmatrix} \\ &= \begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix} \begin{pmatrix} 1-c & d \\ c & 1-d \end{pmatrix} \\ &= \begin{pmatrix} * & d(1-b-a) + b \\ c(1-b-a) + a & * \end{pmatrix}, \end{aligned}$$

which is true if and only if $-bc = -ad$ for all a and b . This can only happen if $c = d = 0$, thus $Z(\mathcal{G}) = \{\mathbf{1}\}$. Now, consider the basic theorem (see for example [5]):

Theorem 1.1. *If a matrix group G is path connected with discrete center, then any non-discrete normal subgroup H will have tangent space $T_1(H) \neq \{0\}$. Further, $T_1(H)$ is an ideal of $T_1(G)$, ie. $[X, Y] \in T_1(H)$ for all $X \in T_1(H)$ and $Y \in T_1(G)$. Therefore, any such H can be detected by checking for ideals of $T_1(G)$.*

In our case, \mathcal{G}^0 satisfies the conditions of this theorem. Suppose \mathcal{I} is a proper ideal of $T_1(\mathcal{G}^0)$. Then \mathcal{I} is one-dimensional, and $Y := xL_1 + yL_2 \in \mathcal{I}$ satisfies:

$$[Y, L_1] = y(L_2 - L_1), \text{ and } [Y, L_2] = x(L_1 - L_2) \in \mathcal{I},$$

which can only be true if $Y \propto (L_1 - L_2)$.

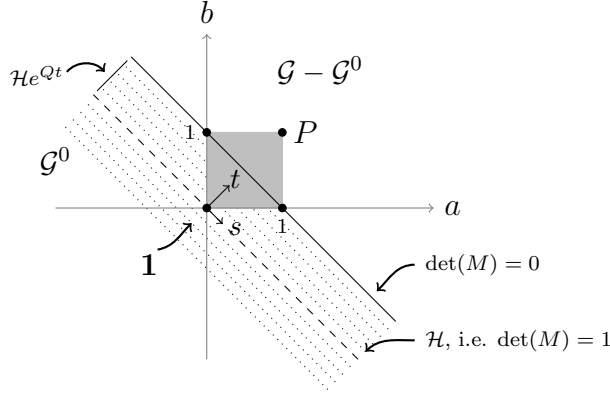


Figure 2: Lie geometry of 2×2 Markov matrices

Result 4. $\langle Y \rangle_{\mathbb{R}} = \langle L_1 - L_2 \rangle_{\mathbb{R}}$ is the only proper ideal of $T_1(\mathcal{G}^0)$.

We take $Y = L_1 - L_2$ and note that $Y^2 = 0$, so $e^{Ys} = e^{(L_1 - L_2)s} = \mathbf{1} + Ys = \begin{pmatrix} 1-s & -s \\ s & 1+s \end{pmatrix} := h_s$. If we define the matrix group $\mathcal{H} := \left\{ \begin{pmatrix} 1-s & -s \\ s & 1+s \end{pmatrix}, s \in \mathbb{R} \right\}$, it is easy to confirm that \mathcal{H} is normal in \mathcal{G}^0 and has tangent space $T_1(\mathcal{H}) = \langle Y \rangle_{\mathbb{R}}$.

Let $\mathbb{R}_{>0}^{\times}$ be the set of positive real numbers considered as a group under multiplication. We have:

Result 5. \mathcal{H} is the kernel of the homomorphism $\mathcal{G}^0 \rightarrow \mathbb{R}_{>0}^{\times}$ defined by $M \mapsto \det(M)$. Thus $\mathcal{G}^0/\mathcal{H} \cong \mathbb{R}_{>0}^{\times}$.

Proof.

$$\det(M) = 1 \Leftrightarrow a + b = 0 \Leftrightarrow M = \begin{pmatrix} 1-a & -a \\ a & 1+a \end{pmatrix}.$$

□

Since $h_s h_t = h_{s+t}$, i.e. \mathcal{H} forms a one-parameter subgroup of \mathcal{G}^0 , we have $\mathcal{H} \cong \mathbb{R}^+$, where $\mathbb{R}^+ = \mathbb{R}$ is considered as a group under addition. Note that $\mathcal{G}^0/\mathcal{H}$ is a parameterised partition of \mathcal{G}^0 , so we can write $\mathcal{G}^0/\mathcal{H} = \cup_{t \in \mathbb{R}} e^{Qt} \mathcal{H}$, where $Q \in T_1(\mathcal{G}^0) - T_1(\mathcal{H})$. We then see that any $M \in \mathcal{G}^0$ can be written as a product $e^{Qt} h_s$, where $\det(M) = \det(e^{Qt})$. Again for reasons of symmetry, we take $Q = \frac{1}{2} \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$, i.e. Q is the generator of the binary-symmetric model, and we have $\det(e^{Qt}) = e^{-t} := \lambda$.

This brings us to our main result.

Result 6. Any $M \in \mathcal{G}^0$ can be expressed as

$$\begin{aligned} M = \begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix} &= e^{Qt} h_s = \frac{1}{2} \begin{pmatrix} 1+e^{-t} & 1-e^{-t} \\ 1-e^{-t} & 1+e^{-t} \end{pmatrix} \begin{pmatrix} 1-s & -s \\ s & 1+s \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 1+\lambda & 1-\lambda \\ 1-\lambda & 1+\lambda \end{pmatrix} \begin{pmatrix} 1-s & -s \\ s & 1+s \end{pmatrix}, \end{aligned} \quad (1)$$

where $\det(M) = \lambda = e^{-t} = 1 - a - b$, and $s = \frac{1}{2}(a - b) \det(M)^{-1}$.

For the binary-symmetric model implemented as a stationary Markov chain, the parameter $\lambda = e^{-t}$ is proportional to the expected number of transitions in chain in time t . Therefore we can think of the parameter s as providing a perturbation away from the binary-symmetric model. Additionally, to ensure that M is a stochastic Markov matrix, with $a, b \geq 0$, we require $-\frac{1}{2}(e^t - 1) \leq s \leq \frac{1}{2}(e^t - 1)$.

The decomposition (1) is the main result of this note and is presented geometrically in Figure 2. It is remarkable that such a simple application of elementary Lie theory has

led directly to this decomposition, and it seems plausible that this decomposition may be useful in practice for (i.) computational efficiency, and/or (ii.) the simple interpretation of the parameters t , λ and s . It will be interesting to explore whether a similar analysis leads to alternative parameterisation for other popular phylogenetic models that form Lie algebras, but we leave this for future work.

References

- [1] J. D. Bashford, P. D. Jarvis, J. G. Sumner, and M. A. Steel. $U(1) \times U(1) \times U(1)$ symmetry of the Kimura 3ST model and phylogenetic branching processes. *J. Phys. A Math. Gen.*, 37:L1–L9, 2004.
- [2] T. House. Lie algebra solution of population models based on time-inhomogeneous markov chains. *J. Appl. Prob.*, 49:472–481, 2012.
- [3] J. E. Johnson. Markov-type Lie groups in $GL(n, R)$. *J. Math. Phys.*, 26:252–257, 1985.
- [4] C. Semple and M. Steel. *Phylogenetics*. Oxford Press, 2003.
- [5] John Stillwell. *Naive Lie Theory*. Undergraduate Texts in Mathematics. Springer, New York, 2008.
- [6] J. G. Sumner, J. Fernández-Sánchez, and P. D. Jarvis. Lie markov models. *J. Theor. Biol.*, 298:16–31, 2012.
- [7] J. G. Sumner, P. D. Jarvis, J. Fernández-Sánchez, B. T. Kaine, Michael D. Woodhams, and B. R. Holland. Is the general time-reversible model bad for molecular phylogenetics? *Syst. Biol.*, 61:1069–1074, 2012.